Student's drinking behavior during on campus/facilities job

Authors: Anand Jain, Charu Saklecha
University of Maryland Baltimore County

Abstract

In the last few years, the usage of wearable devices and how they are tracking the physiological data like Electrodermal Activity(EDA), Heart Rate etc. to improve our health has been extensively researched and used. Similarly, it is necessary to keep an eye on the water intake in our body, but no or very few attempts have been made to research the adverse impact of dehydration on these signals. In this paper, we investigate the changes in physiological signals of body using wearable device. The paper also delineates an empirical evaluation of any classification performance through the development of a machine learning algorithm for automatically detecting EDA and Heart Rate extracted feature artifacts. We compare the water drinking behavior of students working at two different job types, with two subcases: pre-water and post-water intake, i.e. before and after the subject has drank a cup of water while working at a desk and facilities job. We discuss the differences discerned in two specific cases under the respective broader scenarios.

a) Introduction

The water level in body play a crucial role in human behavior. Recent findings suggest that dehydration affects cognitive performance, mental clarity, sleep quality, mood endurance. As hydration status is critical to the body's process of temperature control. Recent findings show that signals like EDA which further derive Skin Conductance Level (SCL) is directly affected from dehydration. The abrupt peaks in phasic conductance, Heart rate verifies the low water intake in the body. Furthermore, prolonged dehydration can lead to stress, anxiety disorders and chronic disease. Therefore, it is important that people are aware of water intake throughout the day. The use of physiological signals to determine the influence of dehydration is relevant as they are originated from the Autonomous Nervous System (ANS) have the advantage - that they cannot be triggered by the user. The approach used in the research paper employs signals such as Electrodermal Activity(EDA), Heart Rate(HR), Inter Beat Interval(IBI). Many research works in the past have used physiological signals to quantify stress, anger and to build emotion recognition model.
The rest of the paper is organized as follows: In section b, we will discuss the Related Work. Our proposed algorithm approach- experimental setup is elucidated in section c. Section d outlines experimental results followed by the Conclusion and Future Work in section e and References following it.

b) Related Work

Physiological signals provide invisible proof of affective arousal [1]. Various studies have been conducted in recent times for understanding the hydration levels and personal hydration needs for an individual. For instance, Vessyl- a smart cup which detects the type of liquid poured into it and displays the caloric content. It helps in regulating caffeine content and managing your weight. The researchers have focused on caloric content of beverages, caffeine and not on the hydration level of body. During the research work, we came across Pryme Vessyl, another smart cup developed by the company as an extension to their previous product Vessyl. It focuses on hydration level and personal hydration needs of an individual by using the dynamic factors like Age, Sex, Activity and weight. In this paper, we use existing wearable technology for collection of physiological data and manual entry of water intake. Previous research work has shown that skin temperature [2] and heart-rate variability [3] are important factors to assess the water level and stress. Our motive is to detect the influence of dehydration on Heart Rate, EDA, in two different scenarios according to the workplace. This can be a part of the solution addressing various industrial use cases such as apps where we just don't feed the amount of water intake but the changes in physiological activity accordingly.

c) Experimental Setup

1. Device Description

To collect the physiological data, we used the Empatica E4 wristband. The wristband tracks collect the data with the help of electrodes and tracks it in real time on an app present on smartphones. The wristband has multiple physiological sensors: BVP, PPG, Temperature, EDA. The BVP signals directly extracts Heart Rate in beats per minute & Inter-Beat Interval from two consecutive beats.

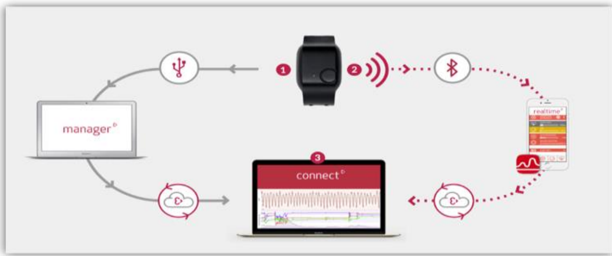

Figure 1- E4 wristband

Sensor	Unit / Range	Sampling Frequency
Temperature	Celsius (°C)	4 Hz
Accelerometer	[-2g, 2g]	32 Hz
Photoplethysmograph		64 Hz
EDA	µS (micro Siemens)	4 Hz
Inter Beat Interval		1/64 second resolution
Heart rate	BPM (beats per minute)	1 Hz

Figure 2-Sensor data specification

2. *Data Collection*

In our experiment, we collected data from 3 different participants, 2 males and 1 female at 2 different workplaces (facilities & desk/sedentary job). The subjects were university students working on campus at different job scenarios, described as follows:

a. When the subject is dehydrated for about 30min-40min, physiological data was collected for 10 minutes and labelled as 'Pre-Water intake'

b. When the subject had a cup of water, after a gap of 30 minutes, physiological data was collected and labelled as 'Post water intake'. The experiment was performed for above mentioned scenarios at a desk job (library) which involved minimum physical activities like typing, minimal hand gestures & facilities (dining hall) which involved activities like lifting, walking, bending & several hand and body related movements.

3. *Methodology & Classification Approach*

a) Data Preprocessing

Data accumulation for every half and hours were not problem-free. For example, participants missed sometimes to time frame for water intake predefined and thereby, had delayed and distorted data recordings.

Firstly, we collect the raw data from Empatica E4 device in individual .csv files for each sensor. Then merging the selected attributes required for the further processing into one single .csv excel file. The input data is formatted into a tabular format where the column values represents the attributes and the row values represents the individual records. We start the preprocessing of the data by first removing records that has one or more missing/null values or any erroneous data for any of the attributes. Like, the EDA, HR data had some noises, missing values which were removed to avoid any discrepancy of data and thus, provide clean, stable dataset. Noise was commonly generated due to the motion of the participants, which caused the disconnection of device and wrist, i.e. for example when there was a gap between the surface of skin and electrodes of the physiological sensors. Further, the raw data collected from the device for different sensors was sampled at different frequencies. So, the next task was the resampling computations so as to bring all the sensors data onto a same frequency to run the different classification algorithms, like specified in figure 2 that EDA data was sampled at 4 Hz, so the EDA data was down sampled to 1 Hz to avoid major loss of data, as during oversampling of Heart Rate which was had very less number of instances as compared to EDA to 2 or 4 Hz.

The Heart rate and EDA were chosen among the attributes for feature extraction and further data analysis.

The feature extraction was done using the heart-rate variability analysis and skin conductance level evaluation using the simple statistical methods, like shown in figure 3.

Feature	Description
Mean HR	Mean of Heart rate(beats per minute)
Standard deviation HR	Standard deviation of instantaneous heart rate values
Mean IBI	Mean of R-R intervals
Standard deviation IBI	Standard deviation of R-R intervals
Mean EDA	Mean of EDA values
Standard deviation EDA	Standard deviation of EDA values

Figure 3- Extracted features

b) Classification Mechanisms-

In this paper, to recognize and differentiate the job type with two subcases: pre-water and post-water intake, we use 4 standard classifiers- i) k-Nearest Neighbour (k-NN), ii) Decision Tree: J48 classifier (DT), iii) Naive Bayes classifier, iv) One Rule classifier. The analysis was done using Weka Tool.

We used the following two approaches. Firstly, the full feature dataset and selected feature dataset was used to check the accuracies and classify the whole dataset into two categories labelled as, facilities and desk job. Secondly, we used only the facilities job type dataset with all the features and then with selected attributes to classify the instances in two class labels, pre-water intake and post water intake and computed the accuracy for both the conditions. The similar process was followed for the desk job-type.

The accelerometer data, x-, y-, z-axis was added to compute precise analysis for detection of job type as desk or facilities, considering the whole dataset.

The accuracy and false positive rate for different classifier models for the whole dataset as obtained is depicted in the figure 4. Then, we select the top 5 features using correlation attribute selection method with ranking algorithm, that came out to be standard deviation of Inter-Beat Interval (IBI), Z-axis parameters, Mean of EDA, standard deviation of EDA and Y-axis values.

The comparison of the accuracies perceived with whole feature set and highest- ranked selected attributes is delineated in following figure 5.

Then we determine the accuracies for Facilities Job for Full Feature set with two class labels, pre-water and post-water intake cases, figure 6. The same process for feature selection was carried out and the top 3 features were chosen- standard deviation of IBI, Mean Heart Rate, Mean IBI. The

accuracies were noted down and compared with full feature facilities job-type dataset, shown in figure 7.
The identical methodology was executed for the desk job-type: with full feature dataset and top 3 ranked selected attributes (Mean IBI, Mean HR, Standard Deviation of EDA) and accuracy for both the situation is depicted in figure 8 and figure 9 respectively.

Classifier	Accuracy	False positive (facility)	False positive (desk job)
Naïve Bayes	96.09993%	0.012	0.059
KNN (K=2)	99.535%	0.005	0.005
OneR	82.8369%	0.134	0.200
Decision Tree	98.8652%	0.008	0.014

Figure 4- Whole dataset

Classifier	Accuracy	Accuracy (Full dataset)
Naïve Bayes	96.09993%	96.1702%
KNN (K=2)	99.535%	99.652%
OneR	82.8369%	82.8369%
Decision Tree	98.8652%	99.007%

Figure 5- Accuracy comparison (whole dataset)

Classifier	Accuracy	False Positive(Without water)	False Positive(With water)
Naïve Bayes	99.8454%	0.003	0.00
KNNN	99.92%	0.00	0.00
OneR	100%	0.00	0.00
Decision Tree	99.9247%	0.002	0.00

Figure 6- Class labeling for Facilities

Classifier	Accuracy	Accuracy(Full Feature)
Naïve Bayes	99.8454%	99.8954%
KNNN	99.92%	100%
OneR	100%	100%
Decision Tree	99.9247%	99.9354%

Figure 7- Accuracy Comparison (Full feature & selected features for Facilities)

Classifier	Accuracy	False Positive(without water	False Positive (with water)
Naïve Bayes	87.1237%	0.014	0.290
KNN	99.4983%	0.003	0.008
OneR	81.9398%	0.194	0.161
Decision Tree	98.8294%	0.009	0.016

Figure 8- Accuracy for Desk Job (Full feature set)

Classifier	Accuracy	Accuracy
Naïve Bayes	87.1237%	74.916%
KNN	99.4983%	94.4816%
OneR	81.9398%	81.9398%
Decision Tree	98.8294%	92.4612%

Figure 9 - Accuracy comparison for Desk Job (Full feature set & selected features)

c) Validation

The raw data observations justified the peaks, abruptions in Heart Rate due to dehydration which are visible in Fig. Similarly, the post-water data had constant Heart Rate with an average of 74.19 beats per minute. The average heart rate was 90.44 beats per minute due to low water level.

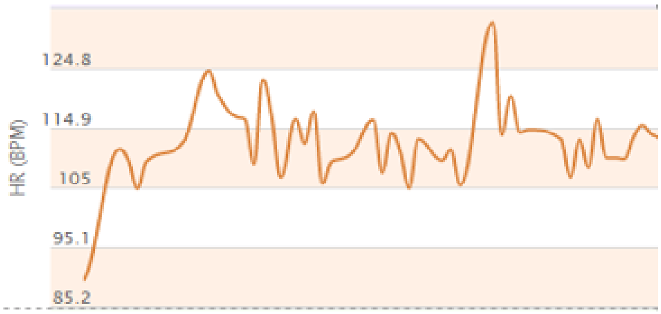

Figure 10- Pre-water (Desk job)

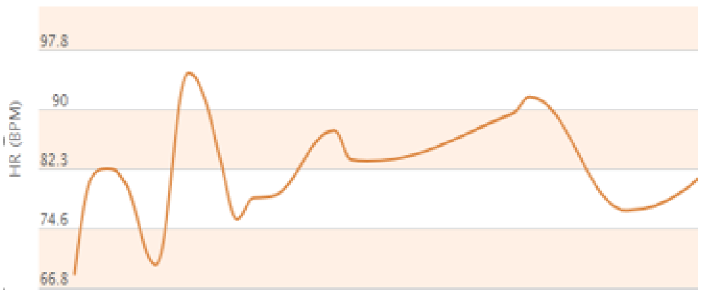

Figure 11- Post-water (Desk job)

d) Experimental Results

The results corroborated the statement that heart rate can be a significant attribute in ascertaining the type of job a person does along with changes in their water intake levels and thereby, explicating the health effects one can pursue in their near future due to these.
It was explicitly depicted from the experimental outcome obtained from operating the trained classification model, on altogether a new test data, unknown to the classifier model, how efficiently and effectively it was able to classify between the kind of job, whether desk (sedentary) or facilities, with categorical sub-class differentiation with-water/without water at different job scenarios and subjects. Out of all the extracted features evaluated from both EDA

and Heart Rate attributes, Standard Deviation of Inter-Beat Interval, IBI was the most dominant feature to distinguish between pre- and post-water cases in a facilities job type, while being the worst feature in desk job. The model is able to classify pre- and post-water efficiently (almost above 90% accuracy for all classifiers) in sedentary position.

e) Conclusion & Future Work

Physiological measures in the real word are complex and to determine the influence of dehydration on them in quite an uphill task. In this paper, we have shown that Heart Rate, EDA are affected significantly due to dehydration. We investigate the striking difference between both the class labels of pre-water intake and Post water intake. Our initial results are promising as classifiers perform and give high accuracies. Our results show that Decision tree is the best classification algorithm. From our work, we can conclude that dehydration influences the physiological signals measured from the user. We performed feature extraction and features selection to improve overall performance.
For a model building perspective, an experiment with a larger dataset embedded with dynamical factors like age and sex can be done. Furthermore, to enhance the performance, additional data should be tracked from wearable device like calories, steps and activity to enhance the scope of research work to a normal individual instead of a university student. It will help to predict the hydration need of a person efficiently. It would be interesting to discover the relationship between the dehydration level and corresponding mood. The amalgamation of physiological data, dynamic factors can be merged with mood recognition models [4] will be helpful to explore the research in this area.

f) References

[1] R. W. Levenson, "Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity," Social Psychophysiology and Emotion: Theory and Clinical Applications, pp. 17–42, 1988.
[2] H. Kataoka, H. Kano, H. Yoshida, A. Saijo, M. Yasuda, and M. Osumi, "Development of a skin temperature measuring system for non-contact stress evaluation," in Engineering in Medicine and Biology Society, IEEE, vol. 2, Oct 1998, pp. 940–943.
[3] G. G. Berntson and J. T. Cacioppo, Heart Rate Variability: Stress and Psychiatric Conditions. Blackwell Publishing, 2007, pp. 57–64.
[4] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "Moodscope: Building a mood sensor from smartphone usage patterns," pp. 389–402, 2013.